Distribution Agreement

Adam Waxman April 17, 2011

What Environmental and Demographic Factors Are Significant Determinants of Obesity?

by

Adam Waxman

Dr. David Frisvold
Adviser

Department of Economics

Dr. David Frisvold

Adviser

Dr. Robert McCauley

Committee Member

Dr. Alessandro Veneziani

Committee Member

April 17, 2011

What Environmental and Demographic Factors Are Significant Determinants of Obesity?

By

Adam Waxman

Dr. David Frisvold

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2011

Abstract

What Environmental and Demographic Factors Are Significant Determinants of Obesity?
By Adam Waxman

In this paper, I investigate potential environmental and demographic factors that may be contributing to the obesity epidemic. I specifically focus on the effects of food away from home (FAFH) expenditure, as well as the effects of restaurant density (both fast food and full-service), on obesity rates in a given area. Using county-level data, my results suggest that FAFH expenditure is positively related to obesity rates. For restaurant density, my model predicts that the number of fast food restaurants per capita is positively correlated with obesity, but that the number of full-service restaurants per capita may be negatively correlated with obesity.

What Environmental and Demographic Factors Are Significant Determinants of Obesity?

By

Adam Waxman

Dr. David Frisvold

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2011

Acknowledgements

This paper was made possible by the support of my adviser, Dr. Frisvold. I'd also like to thank Dr. O'Reilly from the Electronic Data Center for helping me gather and organize the data, as well as Dr. Curran and Dr. Mialon for all their excellent comments and suggestions. Lastly, I would like to thank Dr. McCauley and Dr. Veneziani for taking their time to be part of my committee and review my paper.

Table of Contents

Introduction

The CDC reports that obesity rates among adults in the United States doubled between 1980 and 2000. Today, over 30% of our country's population is obese (defined as body mass index (BMI) greater than 30).[1] While some say that obesity is solely an appearance issue, the increased costs created by this trend suggest otherwise. In terms of health, many studies show being obese increases the risk for serious health concerns such as heart disease, diabetes, and various cancers.[2] As a result, between rising medical costs and a decline in productivity, this epidemic is placing a huge burden on the economy. According to a recent NBER study (Cawley and Meyerhoefer 2010), the medical care costs of obesity-related illness are approximately $186 billion per year or about 16.5% of the country's national health expenditures.

From an economics perspective, there are a few leading hypotheses for the causes of the rapid rise in obesity. First, some economists (i.e. Chou, Grossman, Saffer 2004) suggest that the decline in real food prices may be a cause of obesity. This theory is based on the downward sloping nature of the demand function—as the price of a good goes down, the quantity demanded increases. Another theory suggests that changes in the labor market, specifically rising real wages and more women in the labor force, may account for part of the upward trend in weight levels. In particular, as the opportunity costs associated with staying at home and cooking increase, consumers change their eating habits. Both of these theories help explain another transformation in American eating habits that has been getting lots of attention lately by researchers and policy

[1] "Facts About Obesity in the United States," http://www.cdc.gov/pdf/facts_about_obesity_in_the_united_states.pdf. 08 March 2011.

[2] "NIH Obesity Research," http://www.cdc.gov/pdf/facts_about_obesity_in_the_united_states.pdf. 08 March 2011.

makers—people are eating out more. As a result of technological innovation and agricultural subsidies, the decline in real food price has enabled agribusinesses and restaurant chains (both fast food and full-service) to sell processed food at relatively low prices. Combined with higher opportunity costs of home-cooked meals, it is no surprise that people are consuming more meals away from the home. Do these economic forces, likely leading to an increased consumption of food away from home, cause obesity?

Another hotly debated issue surrounding the obesity conversation is the effects of fast food restaurants on weight outcomes. This topic is closely related to the previous question regarding FAFH. If increased FAFH consumption is a cause of obesity, is the increased availability of energy-dense and low nutritious fast food a contributing factor? What about full-service restaurants?

In this paper I use newly calculated county-level obesity estimates from the Centers for Disease Control, combined with demographic and business data collected from a variety of sources, to examine the impact of different environmental and demographic factors on obesity rates. I specifically am interested in studying the effects of FAFH consumption and restaurant availability (both fast food and full-service) on obesity rates. Adding to the existing literature in this subject is important because understanding the mechanisms and causes behind this increasingly costly issue is vital to promote positive changes through various policies. In this paper I employ the ordinary least squares (OLS) method to determine the individual effects of these different variables. Due to the synthetic nature of a lot of the data, which is explained in more detail in the data section, I employ a cross-sectional analysis of the population.

Background Literature

With so many costs to both individuals and society as a whole, there is much research being done related to potential causes of obesity. In addition to the branches of this subject concerned with genetic and psychological factors that may facilitate weight gain, there is a lot of attention focusing on possible demographic and environmental causes. This increased concern about the environment is partly due to popular books and movies such as Eric Schlosser's *Fast Food Nation* (about the dark side of fast food) and Kelly Brownell's *Food Fight* (which helped coin the term "toxic" food environment). While a lot of the research in the past has been interested in learning possible socioeconomic and ethnic relationships with obesity rates, I am going to limit my literature review to the links that I am most interested in studying: food away from home consumption, full-service restaurants per capita, and fast food restaurants per capita.

Food Away from Home

There are two main papers that have specifically investigated the link between FAFH consumption and obesity. The first study (Binkley, Eales, and Jekanowski 2000) tries to determine if the source from which food is obtained is a significant determinant of obesity. The study uses panel data from the Continuing Survey of Food Intake by Individuals (CSFII) to investigate this topic. This survey is a nationally representative sample of 16,103 individuals, gathering data about what they ate during the past 24 hours on two non-consecutive days. The results suggest that FAFH, especially fast food consumption, is likely to be contributing to the obesity epidemic.

The second and more recent study (Cai, Alviola, Nayga, and Wu 2008) looks at the issue from a more macro standpoint. Similar to the first study, this paper uses panel data. In contrast to the individual responses that the first research paper uses, this study looks at state-level data derived from the Centers for Disease Control's Behavioral Risk Factor Surveillance System (BRFSS). The results suggest that FAFH expenditures are positively correlated with obesity, while food-at-home (FAH) expenditures are negatively related to obesity. Although statistically significant, the magnitudes of these effects are economically relatively small. Specifically, the results indicate that a $1,000 increase in annual per capita FAFH expenditures is associated with a 0.053% increase in obesity rates.

Fast Food and Restaurants

There are many papers that focus specifically on the correlation between fast food and full-service restaurants and obesity. However, the magnitude, and in some cases the sign, of the variables of interest are different depending on the methods and data used.

One of these papers (Jeffery, Baxter, McGuire, and Linde 2006) sets out to look at the relationship between living near fast food restaurants and BMI. They obtained information on 1,033 individuals from Minnesota via a telephone survey. The results suggest that proximity to fast food restaurants is uncorrelated with BMI. However, the study finds that eating at fast food restaurants is positively correlated with BMI.

A 2010 study (Jilcott, MGuirt, Imai, and Evenson) examines the effects of the overall retail food environment on obesity. In order to characterize the food environment, they develop a variable called the Retail Food Environment Index (RFEI). This variable tries to quantify the food environment by taking the number of fast food restaurants and convenience stores and dividing them by the number of supermarkets. To calculate the amount of each of these items, the study uses two sources. First, they use ReferenceUSA business database. They also use the North Carolina Department of Environmental Health billing records. For the dependent variable, the study uses county-level BMI data from the North Carolina Risk Factor Surveillance System survey. As expected, the data suggests that the RFEI is positively correlated with BMI. However, when isolating the different types of food venues, the results indicate a negative correlation between fast food restaurants per capita and BMI.

The previous study mentioned looks at county-level data for obesity figures and more local sources for fast food, convenience stores, and supermarkets. Another recent study (Rashad, Grossman, and Chou 2006) flips around the robustness of the data sources. For BMI and obesity data, this study uses micro-level data from the First, Second, and Third National Health and Nutrition Examination Surveys (NHANES I, II, and III) conducted by the National Center for Health Statistics (NCHS). The per capita number of restaurants (which combines fast food and full-service) is gathered at the state-level from the Census of Retail Trade. The results suggest a positive and statistically significant correlation between restaurants and obesity rates.

Another study (Chou, Grossman, and Saffer 2004) sets out to try and find reasons for the explosion of obesity rates in the U.S. since the late 1970s. This study uses BRFSS from 1984-1999 for state-level BMI estimates. The number of fast food restaurants and full-service restaurants are taken from the Bureau of the Census's 1982, 1987, 1992, and 1997 Census of Retail Trade. In initial regressions, the coefficients of the two types of restaurants were very similar, so they decided to sum the results for their final regression. The results indicate that restaurants (combined fast food and full-service) have a positive and significant correlation to state BMI levels.

The last study (Binkley 2008) adds another interesting spin to the analysis of the impact of fast food and full-service restaurants on obesity. Instead of looking at the number of food venues and the percent of the surrounding population that is obese, this study looks at the calorie intake difference for fast food meals versus full-service restaurants. For data, the study uses the same CSFII survey used by the Binkley, Eales, and Jekanowski (2000) study. The survey gathered the type of food eaten, where it was obtained, and whether it was eaten at breakfast, brunch, lunch, dinner, or as a snack. The results suggest that larger meals are consumed at full-service restaurants, but that these larger meals lead to less consumption at other times of the day, which narrows, and often even reverses this difference in consumption. Although not directly related to what I am looking at, these results illuminate important aspects to think about when evaluating the effects of FAFH consumption on obesity rates.

Contribution

By conducting this study, I hope to contribute to the existing literature by utilizing different data sources to analyze these issues on a national scale using county-level data. The county-level data will allow me to balance the importance of having a large number of observations while also trying to cover a large geographic area. Regarding the two studies that focus on FAFH consumption, one uses a survey of 16,103 individuals, while the other uses state-level data. The first survey has its advantages, as it has exact measures for how much FAFH an individual consumed during two non-consecutive days. However, this survey does not take into account the relative availability of FAFH in different geographic environments. As I am interested in understanding the effects of the so-called “toxic” environment, not having the food venue information is a major drawback. In response to the data in the second study, which uses state-level data for obesity rates, more local levels of data will hopefully help lead to more tailored and efficient policies.

Concerning the studies examining the effects of fast food and full-service restaurants per capita on obesity, I again hope to add to the existing discussion by making use of county-level data sources on a national scale. So far, the studies have either drawn on very detailed information for a small geographic area, or analyzed national trends using state-level data. Similarly, many of the studies have combined fast food and full-service restaurants. As Binkley (2008) demonstrates, we should use caution when evaluating the effects of different food sources away from home. For this reason, I will evaluate these variables separately.

Conceptual and Empirical Model

I conceptualize the obesity function for my study based on Cai et al. (2008). In their study, they use the classic energy balance approach to estimate BMI levels. This approach assumes that the energy balance at time t is the difference between calorie consumption and energy expenditure:

(1) $E_t = C_t - W_t$

In this equation, E_t is the energy balance at time t, C_t is the calorie intake at time t, and W_t is energy expenditure at time t. Accordingly, this equation states that the energy balance of an individual at time t is the cumulative energy balance of all previous time periods. As BMI is directly related to energy balance, this equation shows that a higher BMI can either be attributed to an increased level of C_t, a decreased level of W_t, or both an increased level of C_t and a decreased level of W_t. In addition to energy balance, BMI is also influenced by the various demographic factors that affect an individual's behavior towards consumption and expenditure of energy. With this in mind, I can write BMI as:

(2) $BMI = f(\sum E_t, \mathbf{X}^*)$,

where BMI is a function of cumulative energy balance from all previous time periods and **X***, which is a vector that includes the various exogenous variables that are likely to influence either the amount of energy an individual will choose to consume or the amount of energy an individual will choose to expend. My data includes both indirect variables that influence E_t and demographic variables that are included in **X***. While most of the variables that influence energy balance indirectly impact C_t (household average FAFH expenditure, household average total food expenditure, fast food expenditure per capita, full-service restaurant expenditure per capita, average household income, average household cigarette expenditure, number of grocery stores per capita, number of convenience stores per capita, fast food restaurants per capita, and full-service restaurants per capita), there is also a variable that indirectly affects energy expenditure (recreation and fitness centers per capita).

Also, as my study is aggregated at the county-level, my dependent variable shifts from BMI to percent of a county obese. This will also add another exogenous variable to the **X*** vector (population). The equation to be estimated is then:

(3) $Y_i = f(\sum E_t, \mathbf{X^*})$,

(4) $= f(FAFH_i, F_i, FFexp_i, Rexp_i, I_i, P_i, Ci_i, G_i, C_{i,}, FF_i, R_i, Rc_i, B_i, H_i, A_i, O_i, M_i)$,

where Y_i is the percent of county i that is obese, which is determined by the different environmental and demographic variables that include: average household food expenditure away from home ($FAFH_i$), average household total food expenditure (F_i), fast food expenditure per capita ($FFexp_i$), full-service restaurant expenditure per capita ($Rexp_i$), average household income (I_i), population (P_i), average household cigarette expenditure (Ci_i), number of grocery stores per capita (G_i), number of convenience stores per capita (C_i), number of fast food restaurants per capita (FF_i), number of full-service restaurants per capita (R_i), number of recreation and fitness facilities per capita (Rc_i), percent of population Black (B_i), percent of population Hispanic (H_i), percent of population Asian (A_i), percent of population "other" (O_i), and percent of population male (M_i).

Empirical Framework

My empirical strategy involves using a linear regression model to estimate the *ceteris paribus* effect of each of the independent variables on the dependent variable (obesity rate in a given county). Specifically I use the ordinary least squares (OLS) method to estimate the following model:

(5) $$Y_i = \beta_0 + \beta_1 FAFH_i + \beta_2 F_i + \beta_3 FFexp_i + \beta_4 Rexp_i + \beta_5 I_i + \beta_6 P_i + \beta_7 Ci_i + \beta_8 G + \beta_9 C_{i,} + \beta_{10} FF_i + \beta_{11} R_i + \beta_{12} Rc_i + \beta_{13} B_i + \beta_{14} H_i + \beta_{15} A_i + \beta_{16} O_i + \beta_{17} M_i + \varepsilon_{i.}$$

As the data is cross-sectional, this model uses the observations of all the counties for one year, 2007, to estimate the beta coefficients. Rather than making an analysis focused around the impact of growth of the various variables, this strategy highlights the differences among the separate counties at a single point in time (and how these differences affect Y_i). The last term (ε_i) represents the unobserved errors, which account for the difference between the actually observed results and the estimated outcomes. Because different geographic areas seem to have different effects on the obesity rates of the population in that given area, I run the regressions using state fixed effects. This means that I create a dummy variable for each state (excluding one base state to avoid perfect multicollinearity). Lastly, the first four variables of interest (average household food away from home expenditure, average household total food expenditure, fast food restaurant expenditure per capita, and full-service restaurant expenditure per capita) are regressed separately due to the fact that they are highly correlated.

Data:

Dependent Variable

The dependent variable, age adjusted estimated percent obese, is measured using the definition of body mass index (BMI), which is one of the most commonly used measures to analyze obesity. This guideline is measured as weight in kilograms divided by height in meters squared, and characterizes anyone with a BMI of 30kg/m^2 or higher as obese. Accordingly, the dependent variable is the estimated percent of a county that has a BMI of 30 or higher.

This variable is from the Centers for Disease Control's "U.S. Obesity Trends" study. This county-level dataset was created as a supplement to the existing state-level data, which is collected through the Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS is the world's largest on-going telephone health survey. Established by the Center for Disease Control (CDC) in 1984, the survey collects data from more that 350,000 adults each year, all on whom are persons 18 and older. To help make the state-level survey data more useful, the CDC created a model to estimate county-level prevalence of diabetes and obesity rates. The model uses Bayesian multilevel modeling, and tests these estimates against the largest couple hundred counties that have enough observations to estimate the county-level data without the Bayesian model. In addition, rates are adjusted by calculating specific rates for three age groups: 20-44, 45-64, and 65+. The committee then uses a weighted sum based on the distribution of these three age groups from the 2000 census to calculate the estimated age adjusted obesity rate for a given county.

While having county-level obesity estimates is extremely useful, there are several limitations to this data worth noting. First, as the BRFSS results are self-reported, obesity is likely to be underestimated; this usually happens because people tend to overestimate height and underestimate weight. Accuracy of self-reporting may vary by region. Next, BRFSS only samples households that have landline telephones. This fact tends to overestimate both the average age and income of the population, as wireless-only households tend to be both younger and have a lower income. Lastly, as with any estimation, the county data are more prone to error than studies that do not use estimation. Although a definite drawback, the numerous benefits of being able to study obesity at a more local level help outweigh this fact. One such benefit is the fact that it is easier to identify distributions of smaller populations. For example, if state-level data suggests that Connecticut has the highest average household income, it is hard to understand how this average income differs across the state. However, with county-level data, it is easier to see how the different variables vary between different areas within a state. If the data is more focused (geographically speaking), then the correlations have easier to understand explanatory powers.

Independent Variables

The independent variables are from three main data sources, two of which use multiple sources and interpretive models to collect and estimate county-level data. The first source, SimplyMap, is a data service that Emory subscribes to that enables researchers to create reports with data from multiple sources. The variables from this source include average household food away from home expenditure, average household income, average household cigarette expenditure, number of fast food restaurants per capita, number of full-service restaurants per capita, population, and the race and gender variables.

The household average FAFH expenditure variable includes, “All meals (breakfast and brunch, lunch, dinner and snacks and nonalcoholic beverages) including tips at fast food, take-out, delivery, concession stands, buffet and cafeteria, at full-service restaurants, and at vending machines and mobile vendors. The variable also includes board (including at school), meals as pay, special catered affairs, such as weddings, bar mitzvahs, and confirmations, school lunches, and meals away from home on trips.” SimplyMap estimates this variable using the Bureau of Labor Statistics’s Consumer Expenditure Survey along with the several other national surveys by the U.S. Census Bureau including the Census, Current Population Survey, and American Community Survey. EASI (a professional demographic forecasting and estimation company) then models these surveys to come up with specific county-level data. SimplyMap and EASI use the same surveys and models to estimate the average household cigarette expenditure variable.

For the food venue statistics, SimplyMap uses the US Census Bureau’s County Business Pattern data. Fast food restaurants are defined as limited-service restaurants engaged in providing food services where customers generally order items and pay before eating. In terms of the North American Industry Classification System (NAICS) codes, this variable includes businesses under the 7222 code. In contrast, restaurants are defined as the number of full-service restaurants engaged in providing food services where customers order and are served while seated. These eating-places coincide with the NAICS code 7221.

For the population, race, ethnic and gender variables, SimplyMap uses the Census, Current Population Survey, American Community Survey, and the Housing Unit Estimates. There are four races that make of the population of each county: White, Black, Asian, and "other". In the population survey Hispanic is considered an ethnicity and contains members in all four races. While the "other" category can be thought of in negative terms as the population that is not White, Black, or Asian, it can positively be described as mostly people who are either American Indian, Alaska Natives, Native Hawaiian, or a multiracial group.

Just like the obesity data, while these variables are available at small levels of geography, caution should be taken, as some of the variables (particularly the food away from home expenditure and total food expenditure variables) are somewhat "synthetic" in the sense that they are estimated using a model that uses data gathered at higher levels of geography.

The next major resource used was the Food Environment Atlas, which is a data source provided by the U.S. Department of Agriculture's Economic Research Service to help people gain a spatial overview of a community's ability to access healthy food and the effects of such access. The program assembles statistics from various existing governmental data sources. The variables from this source include the number of grocery stores, the number of convenience stores, and the number of recreation and fitness centers, all at the county-level. The Food Environment Atlas gathers these variables from the U.S. Bureau's County Business Pattern findings. To get a per capita measure, I divided these numbers by the populations from SimplyMap.

Grocery stores are defined to be the number of supermarkets and grocery stores in a county. This variable includes both larger establishments generally known as supermarkets and smaller grocery stores that specialize in a more specific line of food (such as delicatessen type establishments). This category excludes convenience stores both with and without gas stations, and also excludes larger general retail stores that also retail food, such as supercenters and warehouse stores. In terms of NAICS codes, this category includes the 445110 and 4452 codes. Convenience stores are defined as stores or food marts that primarily engage in retailing a limited amount of food goods (i.e. milk, bread, soda, and snacks). This variable includes both stores at gasoline stations and stand-alone businesses. In terms of NAICS codes, this includes the 445120 and 447110 categories.

Lastly, recreation facilities are defined as centers that are primarily engaged in operating fitness and recreational sports facilities, which feature exercise and other conditioning or recreational sports activities. This category is consistent with the NAICS code 713940.

The 2007 Economic Census from the U.S. Census Bureau was the last data source used. This source was used to gather the fast food expenditure and restaurant expenditure data, which was defined as sales for the NAICS codes 7222 (fast food restaurants) and 7221 (full-service restaurants). There was only data available for 1,618 counties for the Fast Food Expenditure variable, and only 626 counties for the Restaurant Expenditure variable. Similar to the Food Environmental Atlas data, to obtain a per capita measure, I divided these numbers by the population from SimplyMap.

Table1 lists the definitions, specific units, and summary statistics for these variables.

Results

Table 2 presents the results from each of the four regressions. On average, the R-squared terms suggest that the independent variables explain just under 80% of the variation in obesity rates between the different counties. Of the seventeen variables of interest, five have positive coefficients that are statistically significant in each regression. These variables include average household FAFH expenditure, average household total food expenditure, number of convenience stores per capita, percentage of the population that is Black, and percentage of the population that is "other". There are six variables that are statistically significant and have negative coefficients in each regression. These variables include average household income, population, average household cigarette expenditure, number of restaurants per capita, number of recreation and fitness centers per capita, and percentage of the population that is Hispanic. In the following paragraphs I will discuss the economic interpretation of each of these variables, talk about the possible reasons for the specific sign, and also consider the limitations of analyzing each of these variables in this specific model.

<u>*Positive Correlations:*</u>

Food Away from Home Expenditure

As shown in regressions (1), the model indicates that FAFH expenditure is positively correlated with obesity. Specifically for every $100 increase in annual average

HH FAFH expenditure for a given population (everything else equal), the model predicts that the percentage of the population that is obese will increase by 1.01%. This coefficient is significant at the 1% level, and is also economically significant; if a population increases its average household food away from home expenditure by $1000 (which is a little more than a third of the average household FAFH expenditure) it is estimated that the obesity rate of the population will increase by about 10%. This result resonates with existing literature about the quality and impact of food away from home on weight outcomes. For example, Lin and Frazao (1997) conclude that meals eaten in restaurants are mostly of lower nutritional quality than meals eaten at home because they usually have higher fat and lower calorie content. This result also agrees with Binkley et al (2000) and Cai et al (2008). However, my result predicts that FAFH expenditure is much more economically significant than the latter study. This correlation is important because it may help explain the impact of the US's trend of increased FAFH consumption over the past several decades. One limitation of analyzing this variable is its likely high correlation with income. Households with more disposable income are likely to spend more on food away from home.

Total Food Expenditure

Similarly, the second regression (2) predicts that average HH food expenditure is statistically significant (at the 1% level) and positive. Economically this coefficient suggests that a $100 increase in average total food expenditure for a population will increase the population's obesity rate by 0.52%. While economically significant, the model predicts that food expenditure has less of an impact compared to FAFH expenditure. This agrees with previous studies that suggest FAFH is less healthy than home-cooked meals. One possible explanation for this relation is that spending more on food might mean that a higher portion of meals are bought at restaurants and fast food locations (which have a higher cost per meal than buying from the grocery store in bulk). Similar to average FAFH expenditure, average total food expenditure is likely highly correlated with income. On average, it is likely that households with higher incomes are going to spend more on food. Also, healthier foods, such as fruits and vegetables, tend to be more expensive than processed food, which suggests a lower value of this coefficient. Lastly, correlation does not necessarily mean causation. It is plausible that obese individuals tend to consume more food than the average person, which causes them to spend more on total food expenditures.

Convenience Stores Per Capita

All of the regressions indicate that the number of convenience stores per capita is significantly and positively correlated with obesity. In the first two regressions, the model predicts that when the number of convenience stores per capita increases by 1000, the obesity rate of the surrounding population will increase by 0.34%. In more understandable terms, this coefficient suggests that for a town of 1,000 people, each additional convenience store will increase the obesity rate by 0.34%. For a town of 100,000 people, adding a convenience store will increase the predicted obesity rate by 0.0034%. While statistically significant, this result is not very economically significant. For example, in a small town of 1,000 people, adding ten convenience stores will only increase the obesity rate of the population by 3.4%. The coefficient of this variable increases from 0.32 and 0.28 in the first two regressions to 1.57 and 4.39 in the third (3) and fourth (4) regressions. One possible reason for this change in magnitude is that there are fewer observations in the last two regressions. Another possible reason is multicollinearity between the number of convenience stores per capita and both the average household expenditure on FAFH and the average household expenditure on total food. The number of observations is likely a better reason as the coefficients change more from the third to fourth regressions then from the first two regressions to the third. The sign and significance of this variable agrees with previous studies on the subject (Jilcott et al. 2010). This result and other studies support the common theory that convenience stores only supply low-nutritious and high caloric food to their customers. An important limitation to this variable it that many convenience stores are located off of exit ramps on interstate highways, indicating that the customers of many convenience stores are not likely local residents. Therefore, it is hard to compare the obesity rates of a given area with the number of convenience stores in that area, as the local population is likely to differ greatly from the customer base of the convenience stores.

Percentage of Population Black and "Other"

The next two variables that are significantly and positively correlated with obesity are racial demographic factors. These two races are Black and "other." As described in the data section, "other" includes those individuals that do not recognize as being White, Black, or Asian. For Black specifically, the model predicts that replacing 1% of a population from White people to African Americans will increase the obesity rate by anywhere from 0.07-0.11%. This result is fairly economically significant, as it says that all other factors equal, an all Black population will have about a 10% higher obesity rate compared to the same population that is all White. The coefficient on Black is consistent with recent studies relating obesity and race (Block et al. 2010). Different cultural eating habits and genetic factors are two possible explanations for this correlation. The "other" variable is harder to discuss as it includes a variety of races and groups. Overall, the positive correlation suggests that genetic and cultural habits of minority races may make them more susceptible to gaining weight.

Fast Food Restaurants Per Capita

Although not statistically significant in all four regressions, the fast food restaurant per capita variable is positive and statistically significantly (at the 1% level for regression 1 and the 5% level for regression 2) in the two regressions with over three thousand observations. This result supports the negative press that suggests fast food restaurants may be a cause of obesity. It also agrees with many recent studies (Chou et al. 2004 and Rashad et al 2006) that predict a positive correlation between fast food restaurants and obesity rates. It is important to recognize the possible endogeneity problem associated with this variable. For example restaurants are not randomly distributed; rather, they choose to be in locations where the demand for their product is higher. If obese people prefer eating in fast food restaurants, then there will be more fast food restaurants near where there are more obese people. This caveat is important because it demonstrates that just because an independent variable is correlated with obesity does not mean that it is necessarily a cause of obesity. Obesity might be a cause of the independent variable or the two variables may randomly be correlated.

Negative Correlations:

Household Income

Average household income is one of five variables in the model that is negatively and significantly correlated with the obesity rate for a given population. Specifically the model predicts that other factors equal, an increase in average household income of $1,000 will decrease the obesity rate by anywhere from about 0.08-0.20%. This may not seem very economically significant, as it would take an increase in average household income of $25,000 to decrease the obesity rate by about 5%. Nonetheless, the negative and significant correlation of income supports previous papers (Chang and Lauderdale 2005) and also endorses possible connections between the prices of healthy food versus high-energy but low nutritious foods. For instance, subsidies of high-energy crops (i.e. corn and soy) and lack of subsidies to more nutritious foods (i.e. fruits and vegetables) have made unhealthy foods more affordable than healthier choices. At the end of the day, though, affordability is only one factor that influences an individual's decision to buy a certain type of food. There are plenty of wealthy people who also buy unhealthy foods.

Population

Population is also negatively and significantly correlated with obesity rates in all four regressions. Specifically the model predicts that an increase in the population by 100,000 people will reduce the obesity rate by anywhere from 0.039-0.062%. This result is statistically significant at the 1% level in all four regressions. However, this result is not very economically significant, as it suggests that an increase of a city's population by 1,000,000 people will only reduce the obesity rate by 0.39-0.62%. The sign and significance of this result is consistent with previous studies (Zhou and Kaestner, 2009) that suggest population density is negatively correlated with obesity. One common explanation is that dense cities tend to have greater access to healthy foods, as there tends to be more grocery stores in highly populated areas compared to rural cities. Also, dense populations tend to be in areas where it is more common to walk for transportation instead driving.

Household Cigarette Expenditure

The average household cigarette expenditure variable is also negatively correlated with obesity in the first two regressions. In these regressions, the variable is statistically significant at the 1% level. In economic terms, this coefficient suggests that for every dollar increase in average household cigarette expenditure, the obesity rate of the local population will decrease by between 0.03-0.04%. This is economically significant, as it proposes that if households spend on average $100 more on cigarettes (which is about a third of the mean), then obesity rates will decrease by 3-4%. In the last two regressions this variable is statistically insignificant. This result is likely due to the fact that average FAFH expenditure and average total food expenditure may be correlated with average cigarette expenditures; an individual who has a higher propensity to consume will consume more of all goods they are interested in.

Recreation and Fitness Facilities Per Capita

Another negatively correlated variable that agrees with previous studies and whose sign is not surprising is recreation and fitness facilities per capita. This variable is statistically significant at the 1% level for the first three regressions, and then statistically insignificant in the fourth (although this regression only has 626 observations and therefore is less robust). As theory suggests, fitness centers likely lead to people being more active, burning more calories, and having a lower BMI. However, the causality could point the other direction. For example, fitness centers will locate near people who demand their services. If people who have low BMIs like to exercise, then fitness centers will locate near them. In this case, the fitness centers would not be causing people to have lower BMIs; instead, fitness centers would be locating near people with low BMIs because that is who likes to use fitness centers.

Full-service Restaurants Per Capita

One of the more interesting results from the model is the fact that the full-service restaurant per capita variable is negatively correlated and significant at the 1% level in all four regressions. This suggests that the more restaurants per capita, the lower the rate of obesity will be. This result contradicts previous studies that suggest that, as the number of restaurants increases, obesity increases (Chou et al. 2004 and Rashad et al 2006). Also, this result weakens the FAFH implication, suggesting that maybe not all FAFH consumption is positively correlated with obesity. The opposing sign for this variable does not necessarily contradict the previous studies, as they both aggregate fast food and full-service restaurants together. Binkley (2008) demonstrates a possible reason for the different effects of various FAFH venues. In his study, he compares the calorie intake of eating at a fast food restaurant versus eating at a full service restaurant. While he finds that people tend to consume more calories during a full-service meal, he also observes that for some reason (maybe hunger or other psychological reasons), people tend to reverse the effect of full-service restaurants by consuming less throughout the rest of the day. While just a single theory, it illustrates the difficulty in pinpointing the causation of different food sources. This non-intuitive result may also highlight the limitations of the model. Whether the fact that the data is very aggregated, or different selection and endogeneity problems, the coefficients of the different variables may be slightly skewed.

Percentage of Population Hispanic

Another interesting result is the coefficient on the percent of a population that is Hispanic variable, which is negative and statistically significant at the 1 percent level in three regressions and significant at the 5% level in the fourth regression. This implication goes against a lot of past studies that suggest Hispanic populations are positively correlated with obesity. For example, Cai et al. 2008, which is a very similar study with mostly the same results, predicts that Hispanic populations are significant and positively correlated with obesity. One possible reason for this difference is the fact that the Cai et al. 2008 study includes several variables that I do not include (i.e. education variables and age variables). Accordingly, my model may leave out an important factor that is highly correlated with obesity rates and the percentage of a population that is Hispanic. Another possible reason is the high correlation (0.85) between the percentage of a population that is Hispanic variable and the percentage of a population that is "other" variable.

Statistically Insignificant Results:

Lastly, the result that fast food and full-service expenditures per capita are uncorrelated with obesity rates is fairly surprising. As fast food expenditure seems like a more direct link to fast food consumption, most people would predict that fast food and restaurant expenditures would be positive and fairly significant. This relation goes against the claim that we live in a "toxic" environment. Instead, it suggests that other factors are more significant determinants of obesity than fast food and full-service restaurant consumption.

Conclusion, Limitations, and Future Research

Overall, the model gives further reason to believe that FAFH consumption is positively correlated with obesity. The study also supports claims that the presence of convenience stores, the average total food expenditure, and the percentage of a population that is either Black or "other," are all positively correlated with obesity rates. Lastly, the study suggests that income levels, population density, recreation centers, full-service restaurants, and percentage of a population that is Hispanic are negatively correlated with obesity.

These results have several implications for how we should try to combat the obesity epidemic. First, the FAFH correlation suggests that efforts should be made to try and give people more incentives to cook-more and eat out less. Whether figuring out a way to decrease the opportunity cost of preparing food at home or increasing the price of FAFH through taxes, there seems to be evidence that reversing the trend of eating out more could benefit both our health and our wallets. A recent push to require restaurants (both fast food and full-service) to post calorie information on their menus is a current real-world example of a type of policy that this research supports. The convenience store and income results suggest that the new focus on trying to understand and eliminate food deserts is right on target. The positive correlation between convenience stores and obesity rates, combined with the negative correlation between income and obesity, suggests policy efforts should be focused on making healthy food both available and affordable.

While agreeing with past studies on many issues, the data and model also highlight several relationships that have not been presented before. These results include the predicted negative correlation between restaurant density and obesity rates and the negative correlation between Hispanic populations and obesity rates. The negative correlation between restaurant density per capita and obesity rates highlights one of the main insights of using this data. By separating fast food from full-service restaurants I am able to uncover an important possible difference between these two types of restaurants and how they affect obesity rates. Nonetheless, extreme caution should be taken before concluding a causal relationship between these two variables. In regards to the full-service restaurant result, there is likely an endogeneity problem. For example, sit-down restaurants (which are more expensive), likely target communities with higher incomes, which tend to be less obese. Looking at the Hispanic population result, the coefficient is likely skewed due to the high correlation between the percentage of a population that is Hispanic and the percentage of a population that is "other" (multicollinearity problem).

Limitations

Lastly, it is important to mention the wide varieties of limitations that the nature of this study presents. First, being aggregated at the county level makes it very hard to understand the mean and distribution of the different variables. For example, 95% of a population in a given county may have income below the poverty rate. However, if the other 5% of the county consists of some of the wealthiest Americans, the average will be very skewed and not represent the true mean. This in turn makes it hard to understand the precise meaning of the various correlations. Furthermore, the results are solely based on associations and not causal links. This limitation is made worse by the nature of this topic. First, there could easily be selection bias in many of the variables. For example healthy people may choose to live in healthy areas near other healthy people. If this is the case, the negative coefficients may not hold as much value as if people were randomly distributed over all of the counties. Secondly, as mentioned throughout the results section, a lot of the variables may be correlated with error terms because the model does not include all the related variables that explain the obesity rate of a certain area. Fast food may cause obesity, but fast food restaurants may also choose to be in areas where there are higher rates of obesity (endogeneity problem). Another limitation of the estimated and high-level dataset is that several of the variables are highly correlated with each other (i.e. the percentage of a population that is Hispanic and the percentage of a population that is “other”). Although this does not affect the explanatory power of the model as a whole, it does affect the preciseness of the individual coefficients.

Future Research Suggestions

The county-level data, along with separating the effects of fast food and full-service restaurants, helps contribute to the task of trying to understand the effects of different environmental and demographic factors on obesity rates. However, these results are only a foundation that must be built upon to further understand these complex relationships. There are several main areas of future research that would supplement these results and help uncover with more confidence possible environmental and demographic determinants of obesity. First, research must be done on the supply side of many of these variables. For instance, for fast food restaurants per capita and full-service restaurants per capita, I only examine the factors that potentially affect the consumer *demand*. In order to fully understand the effect of food venues, a study must try and build a model to predict how *supply* factors fit into the overall equation. Understanding how fast food and full-service restaurants locate is important to fully understand the causal relationship between these variables and obesity rates. Next, the results of my study suggest that there should be added concentration on trying to understand the different effects of fast food restaurants and full-service restaurants on obesity. In addition to adding the supply-side model into the equation, more focus should be put on specific and direct effects of eating at fast food restaurants versus full-service restaurants (similar to Binkley, 2008). Lastly, the positive and statistically significant coefficient on the food away from home expenditure variable suggests that future research should focus on possible reasons for this trend. These research topics include the effects of women in the labor force, shifts in cooking education, and any other topics that increase the costs of preparing and eating meals at home.

References

Amarasinghe, A., C. Brown, and G. D'Souza. "The Impact of County Level Factors on Obesity in West Virginia." Working Paper.

Binkley, J.K. "The Effect of Demographic, Economic, and Nutrition Factors on the Frequency of Food Away from Home." *The Journal of Consumer Affairs* 40(2006):372-391.

Binkley, J.K. "Calorie and Gram Differences Between Meals at Fast Food and Table Service Restaurants." *Review of Agricultural Economics* 30(2008):750-763.

Binkley, J.K., J. Eales, and M. Jekanowski. "The Relation Between Dietary Change and Rising US Obesity." *International Journal of Obesity* 24(2000):1032-1039.

Block, J., R. Scribner, and K. DeSalvo. "Fast Food, Race/Ethnicity, and Income: A Geographic Analysis." *American Journal of Preventive Medicine* 27(2004):211-217.

Cai, Y., P. Alviola IV, R. Nayga Jr., and X. Wu. "The Effect of Food-Away-from-Home and Food-at-Home Expenditures on Obesity Rates: A State-Level Analysis." *Journal of Agricultural and Applied Economics* 40(2008):507-521.

Cawley, J. and C. Meyerhoefer. "The Medical Care Costs of Obesity: An Instrumental Variables Approach." Working Paper.

Centers for Disease Control and Prevention, Behavioral Risk Factor Surveillance System Survey Data, Atlanta: U.S. Department of Health and Human Services, Internet site: www.cdc.gove/brfss.

Chang, V., and D. Lauderdale. "Income Disparities in Body Mass Index and Obesity in the United States, 1971-2002." *Arch Intern Med* 165(2005):2122-2128.

Chou, S.Y., M. Grossman, and H. Saffer. "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23(2004):565-587.

Currie, J., S. Della Vigna, E. Moretti, and V. Panthania. "The Effect of Fast Food Restaurants on Obesity and Weight Gain." Working Paper.

Flegel, K.M., M. Carroll, C. Ogden, and L. Curtin. "Prevalence and Trends in Obesity Among US Adults, 1999-2008." *Journal of the American Medical Association* 303(2010):235-241.

Gillis, L., and O. Bar-Or. "Food Away from Home, Sugar-Sweetened Drink Consumption and Juvenile Obesity." *Journal of the American College of Nutrition* 22(2003):539-545.

Gruber, J. and M. Frakes. "Does Falling Smoking Lead to Rising Obesity?" *Journal of Health Economics* 25(2006):183-197.

Guthrie, J., B. Lin, and E. Frazao. "Role of Food Prepared Away from Home in the American Diet, 1977-78 versus 1994-96: Changes and Consequences." *Journal of Nutrition Education and Behavior* 34(2002):140-150.

Jeffery, R., J. Baxter, M. McGuire, and J. Linde. "Are Fast Food Restaurants an Environmental Risk Factor for Obesity?" *International Journal of Behavioral Nutrition and Physical Activity* 3(2006).

Jekanowski, M., J. Binkley, and J. Eales. "Convenience, Accessibility, and the Demand for Fast Food." *Journal of Agricultural and Resource Economics* 26(2001):58-74.

Jilcott, S.B., J. McGuirt, S. Imai, and K. Evenson. "Measuring the Retail Food Environment in Rural and Urban North Carolina Counties." *Journal of Public Health Management and Practice* 16(2010):432-440.

Lee, J., and M. Brown. "Food Expenditures at Home and Away From Home in the Unites States: A Switching Regression Analysis." *The Review of Economics and Statistics* 68(1986):142-147.

Lin, B. and E. Frazao. "Nutritional Quality of Foods at and Away from Home." *Food Review* 2(1997): 33-40.

Morland, K., A. Diex Roux, and S. Wing. "Supermarkets, Other Food Stores, and Obesity: The Atherosclerosis Risk in Communities." *American Journal of Preventive Medicine* 30(2006):333-339.

Rashad, I., M. Grossman, and S. Chou. "The Super Size of America: An Economic Estimation of Body Mass Index and Obesity in Adults." *Eastern Economic Journal* 32(2006):133-148.

Raynor, H., and L. Epstein. "Dietary Variety, Energy Regulation, and Obesity." *Psychological Bulletin* 127(2001):325-341.

Zhao, Z., and R. Kaestner. "Effects of Urban Sprawl on Obesity." Working Paper.

Table 1
Definitions, Means, and Standard Deviations of Variables (2007)

Variable	Definition	Mean (Standard Deviation)
Obesity	% Population obese	28.25
		(3.69)
FAFH Expenditure	Food expenditure away from home (HH avg $00)	25.51
		(1.72)
Total Food Expenditure	Food expenditure (HH avg $00)	58.37
		(2.94)
Fast Food Expenditure / Capita	Fast food expenditure per capita ($00)	5.01
		(2.97)
Restaurant Expenditure / Capita	Full-service restaurant expenditure per capita ($00)	7.08
		(6.05)
Income	Income (HH avg $000)	52.88
		(12.36)
Population	# of people (x100,000)	0.96
		(3.11)
Cigarette Expenditure	Cigarette expenditure (HH avg $)	289.47
		(6.65)
Grocery Stores / Capita	# Grocery stores per capita (x1,000)	0.29
		(0.23)
Convenience Stores / Capita	# Convenient stores per capita (x1,000)	0.64
		(0.32)
Fast Food Restaurants / Capita	# Fast food restaurants per capita (x1,000)	0.69
		(0.42)
Full-service Restaurants / Capita	# Full-service restaurants per capita (x1,000)	0.78
		(0.63)
Recreation and Fitness Centers / Capita	# Recreation and fitness facilities per capita (x1,000)	0.85
		(0.84)
Black	% Population Black	9.04
		(14.97)
Hispanic	% Population Hispanic	7.51
		(13.28)
Asian	% Population Asian	1.06
		(2.83)
Other	% Population Other	7.13
		(10.59)
Male	% Population Male	49.87
		(1.90)

Table 2
County-Level Linear Regression with State Fixed Effects (2007)

	1	2	3	4
	Age Adjusted Percent Obese			
FAFH Expenditure	1.013***	--	--	--
	(0.073)			
Total Food Expenditure	--	0.520***	--	--
		(0.035)		
Fast Food Expenditure / Capita	--	--	0.028	--
			(0.019)	
Restaurant Expenditure / Capita	--	--	--	0.031
				(0.022)
Income	-0.202***	-0.189***	-0.075***	-0.085***
	(0.009)	(0.008)	(0.006)	(0.009)
Population	-0.062***	-0.061***	-0.061***	-0.039***
	(0.012)	(0.012)	(0.013)	(0.015)
Cigarette Expenditure	-0.041***	-0.031***	0.007	-0.002
	(0.009)	(0.008)	(0.013)	(0.021)
Grocery Stores / Capita	-0.196	-0.223	-0.315	-2.692*
	(0.158)	(0.157)	(0.666)	(1.442)
Convenience Stores / Capita	0.320***	0.297***	1.573***	4.393***
	(0.115)	(0.115)	(0.338)	(0.869)
Fast Food Restaurants / Capita	0.242**	0.291***	0.2	-0.716
	(0.096)	(0.096)	(0.289)	(0.611)
Full-service Restaurants / Capita	-0.474***	-0.480***	-1.867***	-1.956***
	(0.071)	(0.071)	(0.192)	(0.386)
Recreation and Fitness Centers / Capita	-1.635***	-1.533***	-3.897***	-2.641
	(0.376)	(0.376)	(1.032)	(1.951)
Black	0.107***	0.109***	0.081***	0.066***
	(0.003)	(0.003)	(0.005)	(0.010)
Hispanic	-0.035***	-0.036***	-0.042***	-0.032**
	(0.004)	(0.004)	(0.007)	(0.015)
Asian	0.01	0.023	-0.114***	-0.065**
	(0.021)	(0.021)	(0.027)	(0.032)
Other	0.078***	0.076***	0.079***	0.086***
	(0.005)	(0.005)	(0.012)	(0.031)
Male	0.025	0.02	0.072*	0.162*
	(0.018)	(0.018)	(0.038)	(0.095)
Constant	22.694***	14.740***	26.351***	24.430***
	(2.101)	(2.174)	(3.670)	(6.115)
Obs	3,140	3,140	1,618	626
R-Squared	0.792	0.793	0.776	0.793

*** p<0.01, **p<0.05, *p<0.1
Note: Standard errors are located in parentheses.